

PROGRAMME

VII^e colloque jeunes chercheurs Praxiling



Corpus, données, modèles: Approches qualitatives et quantitatives

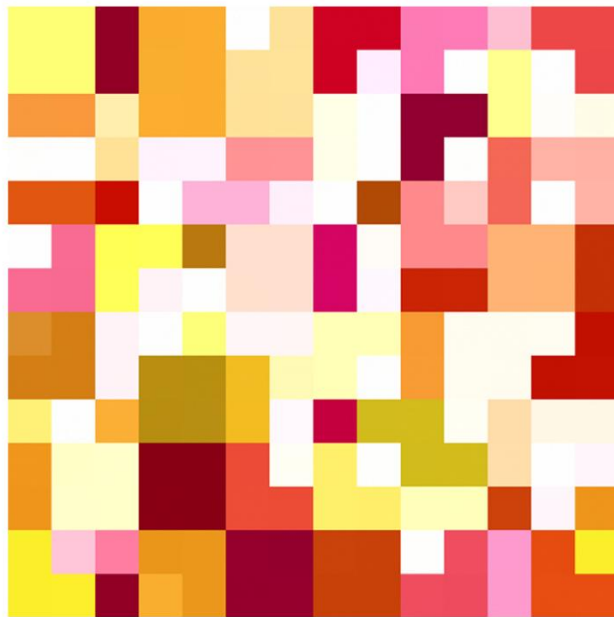


image CC by 2.0 par Patrick Hoesly

9 et 10 Juin

Université Paul-Valéry Montpellier 3 - Bâtiment D

Organisation: Lucie Alidières, Sophie Azzopardi, Camille Capelle,
Hyeran Lee, Véronique Carrière et Clair-Antoine Veyrier

<http://praxiling.univ-montp3.fr>

Pour en savoir plus:



ED58 LANGUES, LITTÉRATURES, CULTURES, CIVILISATIONS
Praxiling UMR 5267 (Université Montpellier 3 - CNRS)

Sommaire

Introduction.....	2
Programme	3
Présentation des conférences.....	Erreur ! Signet non défini.
Jeudi 9 Juin.....	Erreur ! Signet non défini.
Vendredi 10 juin.....	Erreur ! Signet non défini.
Communications affichées.....	Erreur ! Signet non défini.
Index des communicants	Erreur ! Signet non défini.

Introduction

La notion de corpus est relativement consensuelle en Sciences du Langage. Il n'est pas rare de considérer que la constitution d'observables linguistiques à partir de données empiriques issues de l'usage est un préalable fondamental à toute théorisation sur la langue. L'ensemble des données recueillies par le chercheur est généralement défini comme « l'univers du discours ». Le corpus se définit quant à lui comme l'ensemble des données recueillies soumises à l'analyse.

Si la majorité des linguistes s'accorde sur le caractère nécessaire de la constitution d'un corpus comme base de toute théorisation, un certain nombre de problèmes se pose inévitablement.

C'est pourquoi la nouvelle édition 2011 du colloque Jeunes Chercheurs Praxiling souhaite appréhender les différentes approches du corpus, interroger les méthodes de constitution et de traitement de corpus ainsi que sa place et sa pertinence dans l'analyse.

Du point de vue méthodologique, en amont, on peut constater qu'il existe d'importantes différences dans les pratiques de constitution de corpus et dans leur traitement. Les diverses conceptions de la notion de corpus renvoient à des domaines hétérogènes en Sciences du Langage qui font appel à des méthodes de traitement variées.

On peut ainsi se demander dans quelle mesure un corpus peut être exhaustif, comment se sélectionne et se délimite un corpus, comment il se constitue et quelle est la nature des données (recueil des données audio, vidéo, numérique, notes, observations ethnographiques). Peut-on et doit-on fixer un seuil minimum du nombre d'occurrences permettant de considérer qu'un corpus est pertinent ? Quelles méthodes sont mises en œuvre pour recueillir ces données ? Comment motiver ses choix pour élaborer un corpus en vue d'une analyse multimodale (corpus vidéos et corpus d'interactions numériques) ? Comment prendre en compte les problèmes éthiques dans le recueil des données ?

Du point de vue scientifique, un corpus pourrait se définir comme une sélection aléatoire d'un ensemble d'énoncés ayant pour but d'être représentatif d'un état de langue à un moment donné. Or, la recherche scientifique s'inscrit dans une démarche se voulant objective et basée sur des données irréfutables. Par conséquent, il semble légitime de se demander en quoi un corpus est représentatif. Dans quelle mesure peut-il être considéré comme valide et pertinent dans le cadre d'une théorisation ? L'hétérogénéité de ses éléments tout comme la subjectivité inhérente aux choix opérés par le chercheur ne constituent-elles pas un obstacle à la démarche scientifique ? Le corpus a été présenté ici comme un composant essentiel de la recherche. On peut cependant s'interroger sur la place qui est accordée au corpus dans l'analyse des données : quelle différence peut-on établir entre les analyses fondées sur corpus (corpus-based) et les analyses guidées par corpus (corpus-driven) ?

Enfin, dans une perspective allant au-delà de la recherche particulière, on peut questionner le lien qui existe entre un corpus et l'objet de recherche : peut-on prétendre élaborer des corpus de référence permettant une pluralité de recherches ou bien n'est-il valide que pour un projet précis à un moment donné ?

Programme
Jeudi 9 juin 2011

8h30-9h00 Accueil des participants - Installation des communications affichées	
9h00-9h30	Ouverture du colloque Bruno Bonu , Directeur du Laboratoire Praxiling, Chantal Charnet , Directrice de l'ITIC
9h30-10h30	Conférencier invité : Damon Mayaffre Corpus et web-corpus. Réflexion sur la « corporalité » à l'heure du numérique Présenté par Bruno Bonu
10h30-11h00 Pause café	
Séance présidée par Jacques Bres : EXPLOITATION DE CORPUS ECRITS	
11h00-11h30	Laurence Longo Un corpus de genres textuels variés pour l'identification automatique des chaînes de référence
11h30-12h00	Addeo Felice et Diana Paolo "Telling Stories": a mixed approach in collecting and analyzing data on values
12h00-14h00 Pause déjeuner	
Séance présidée par Jacques Bres : ENJEUX AUTOUR DE LA CONSTITUTION D'UN CORPUS DE REFERENCE	
14h00-14h30	Catherine Pinon Quel corpus pour fonder la grammaire d'une langue pluriglossique ? Exemple de l'arabe standard contemporain.
14h30-15h00	Simon Landron, Jiayin Gao, Nikola Paillereau, Ahmad Nawafleh, Christelle Exare et Hirofumi Ando Vers la construction d'un corpus commun de français langue étrangère : pour une étude phonétique des productions de locuteurs de langues maternelles plurielles.
15h00-15h30	Stéphanie Lopez Corpus de référence et corpus d'usages : méthodologie de constitution pour une analyse des communications pilote-contrôleur
15h30-16h00	Guri Bordal L'étude prosodique de la parole spontanée: le cas du français centrafricain.
16h00-16h30 Pause café	
Séance présidée par Fabrice Hirsch : CONSTITUTION / TRAITEMENT DE CORPUS VIDEO	
16h30-17h00	Vassiliki Markaki Filmer le désaccord en réunion : réflexions sur l'enregistrement vidéo de phénomènes interactionnels complexes
17h00-17h30	Clémentine Hugol-Gential L'utilisation d'un restaurant expérimental pour la captation de données audio et vidéo : questions et enjeux
17h30-18h00	Isabel Colon de Carvajal Du corpus enregistré au corpus analysé : l'utilisation d'outils de requêtes informatisés
18h00-18h15	Présentation de Corpus Pluriels par Chantal Charnet et Camille Capelle
18h15-20h00 Présentation des communications affichées autour d'un verre	

Vendredi 10 juin 2011

8h30-9h00	Accueil
9h00-10h00	Conférencier invitée : Agnès Witko Corpus et pathologie du langage. Quelle approche en Orthophonie-Logopédie ? Présentée par Fabrice Hirsch
PARTAGEABILITE, EXPLOITATIONS PLURIDISCIPLINAIRES D'UN MEME CORPUS Séances présidées par Damon Mayaffre	
10h00-10h30	Delphine Giuliani et Radia Hannachi IntUne, constitution et exploitations d'un corpus.
10h30-11h00	Aude Wirth Le Corpus de documents comptables médiévaux (CDCM) : genèse, caractéristiques et intérêts.
11h00-11h30	Pause café
11h30-12h00	Sophie Dalle-Nazébi, F. Lefebvre-Albaret, J. Dalle et P. Dalle Traitement des corpus en langue des signes en recherches socio-linguistiques
12h00-12h30	Alexandra Caria Informen ethnographiquement le chercheur en analyse de conversation : le cas de la data session
12h30-13h	Lolita Berard et Christophe Benzitoun Mutualisation de corpus oraux : retour d'expérience
12h30-14h30	Pause déjeuner
14h30-15h30	Conférencier invitée : Lorenza Mondada Préservation et traitement de la temporalité dans les corpus de parole en interaction Présentée par Bruno Bonu
	ASPECTS PROBLEMATIQUES DE LA TRANSCRIPTION : TEMPORALITE, GESTUALITE, ORALITE Séances présidées par Bruno Bonu
15h30-16h00	Samira Moukrim Sur la constitution de corpus de deux langues à tradition orale (le berbère tamazight et l'arabe marocain) parlées à Orléans
16h00-16h30	Raphaël de Courville, Claire Danet, Morgane Rébulard et Roman Miletitch Qu'est ce qu'un corpus tracé de locuteurs/scripteurs de Langue des Signes Française (LSF) en vue de la constitution d'une écriture ?
16h30-17h00	Geneviève Pinard-Prévoist La transcription perceptuelle au service du corpus oral en contexte naturel
17h00-17h30	Clôture du colloque

Communications affichées

Ibrahima Abdoul H. Cissé (GIPSA-lab (Grenoble, France) LUCL (Leiden, Pays-Bas))

Constitution d'un large corpus audio-vidéo en acquisition du langage chez des bébés âgés de 6 à 38 mois en terre africaine : défis techniques, sociolinguistiques et éthiques

Géraldine Charles-Dominique (Cerca-CNRS, Université de Poitiers)

La compréhension de documents multiples en langue 1 et 2 : l'analyse de corpus naturel pour tester les effets d'un outil pédagogique.

Anne Küppers (VALIBEL - Discours et Variation, Université catholique de Louvain),

Lydia-Mai HO-DAC, (CLLE-ERSS, Université de Toulouse-Le-Mirail),

Un corpus de presse francophone pour l'étude de l'impact d'Internet sur les pratiques langagières

Marion Mauger-Parat (CERLIS, Paris Descartes)

Changement climatique et débats d'idées. Représenter le processus de transmission d'un concept scientifique controversé.

Frédéric Mazières (Université Paris III)

L'enquête sociolinguistique et la diffusion du français

Katérina Palasis (BCL UMR 6039 CNRS, Université de Nice Sophia-Antipolis)

Pour une meilleure représentativité des données orales en grammaire générative : problèmes et solutions à partir d'un nouveau corpus enfantin français

Mirsini Tzanavari (Université de Macédoine)

Différentes méthodes d'exploration textuelle outillées appliquées à un même corpus

Emmanuel Quintin (Paris-Sorbonne 4)

Vers une détection et une caractérisation automatique des entités nommées : étude fondée et orientée sur corpus d'antonomasés

Kadi Oum-El-Djilali (Université de Mostaganem)

L'analyse linguistique de corpus spécialisés à des fins didactiques

Comité scientifique :

Bruno BONU (Praxiling UMR 5267 Université Montpellier 3-CNRS)
Melissa BARKAT-DEFRADAS (Praxiling UMR 5267 Université Montpellier 3-CNRS)
Olivier BAUDE (Coral EA 3850, Université d'Orléans)
Roxanne BERTRAND (Parole et Langage - UMR 6057 CNRS, Université de Provence)
Mireille BILGER (CILSH, Université de Perpignan)
Jacques BRES (Praxiling UMR 5267 Université Montpellier 3-CNRS)
Mathias BROTH (Institutionen för kultur och kommunikation (IKK), Linköpings Universitet, Suède)
Anne DISTER (Centre de recherche VALIBEL / CENTAL, Université catholique de Louvain)
Sylvain KAHANE (MoDyCo UMR 7114 CNRS, Université Paris X)
Sarah LEROY (MoDyCo UMR 7114 CNRS, Université Paris X)
Damon MAYAFFRE (UMR 6039 BCL – CNRS, Université de Nice Sophia-Antipolis)
Sylvie MELLET (UMR 6039 BCL – CNRS, Université de Nice Sophia-Antipolis)
Lorenza MONDADA (ICAR UMR 5191, Université Lyon 2)
Chrysta PELLISIER (Praxiling UMR 5267 Université Montpellier 3-CNRS)
François RASTIER (Ertim-INALCO, Université Paris X)

Comité d'organisation :

Lucie ALIDIÈRES
Sophie AZZOPARDI
Camille CAPELLE
Véronique CARRIERE
Hye Ran LEE
Clair-Antoine VEYRIER



Laboratoire Praxiling UMR 5267
(Université Montpellier 3 –CNRS)
<http://praxiling.univ-montp3.fr>

Pour suivre toute l'actualité de Praxiling et du CJC 2011, abonnez-vous au fil Twitter :
<http://twitter.com/Praxiling>